

POLITIQUES &
MANAGEMENT
PUBLIC

Politiques et management public

Vol 31/1 | 2014

L'évaluation des politiques publiques

Les sources intellectuelles de l'évaluation des politiques publiques

The intellectual origins of programme evaluation

Jean-Claude Barbier



Electronic version

URL: <http://journals.openedition.org/pmp/6947>

ISSN: 2119-4831

Publisher

Institut de Management Public (IDPM)

Printed version

Date of publication: 1 January 2014

Number of pages: 13-30

ISBN: 978-2-7430-1995-2

ISSN: 0758-1726

Electronic reference

Jean-Claude Barbier, « Les sources intellectuelles de l'évaluation des politiques publiques », *Politiques et management public* [En ligne], Vol 31/1 | 2014, mis en ligne le 10 décembre 2016, consulté le 01 mai 2019. URL : <http://journals.openedition.org/pmp/6947>

Les sources intellectuelles de l'évaluation des politiques publiques

01

► **Jean-Claude Barbier**

CNRS CES, Université Paris 1 Panthéon Sorbonne

Résumé

L'évaluation des politiques publiques est une « discipline » délicate à situer dans le paysage du management public. Routinisée dans certains pays où elle dispose de références professionnelles, elle est moins fermement identifiée dans d'autres. Dans les deux cas cependant, elle s'insère dans une famille d'activités qui traitent d'information évaluative. Le contrôle de gestion est une discipline voisine. On comprend mieux la cohérence de l'évaluation de politiques publiques en remontant à ses sources intellectuelles. Historiquement, la première est la forme d'expérimentalisme développée par Donald Campbell, dont il reste dans l'évaluation pratiquée aujourd'hui comme une nostalgie scientifique. Mais le management public, et en particulier sa version dite « nouvelle gestion publique » constituent le milieu d'accueil de la discipline. Comme le management, l'évaluation allie un « état-de-l'art » transnational à des pratiques nationales diversifiées.

© 2014 IDMP/Lavoisier SAS. Tous droits réservés

Mots clés : expérimentalisme, évaluation, politiques publiques, management public.

Abstract

The intellectual origins of programme evaluation. Programme evaluation is not evenly identifiable across nations, where it is always associated with public administration and management. In many countries, programme evaluation has accumulated a considerable amount of practice and reflection, whereas in others achievements have remained more ambiguous. In any case, the latter are always accomplished in the context of multiple connections to sister disciplines as performance audit, quality management and reporting. Exploring the intellectual origins of programme evaluation allows for a better understanding of its content. Historically, experimentalism, initially introduced by Donald Campbell, was essential in the foundation of programme evaluation. But public management is its natural milieu, including its recent variation, New Public Management. Just like manage-

*Auteur correspondant : Jean-Claude.Barbier@univ-paris1.fr

doi :10.3166/pmp.31.13-30 © 2014 IDMP/Lavoisier SAS. Tous droits réservés

ment, programme evaluation has two components : a transnational “state-of-the-art” and a great variety of national references.

© 2014 IDMP/Lavoisier SAS. Tous droits réservés

Keywords: experimentalism, programme evaluation, policy analysis, public administration, management.

Introduction

Parcourant la distance historique entre la fin des années 1990 et aujourd’hui, on ne peut manquer d’être frappé par l’écart existant entre les représentations publiques de l’évaluation de l’époque et celles de 2014. Certes le mot évaluation a toujours été hautement polysémique. Chacun d’entre nous évalue dans la vie ordinaire, et le sait parfaitement. Mais là n’est pas l’important. À la fin des années 1990, quand existait un Conseil scientifique de l’évaluation (CSE) et que se fondait pour la première fois en France une « Société française de l’évaluation¹ », au sein des « forums de communautés de politiques publiques² » l’évaluation signifiait l’une des voies de « modernisation » de l’action publique (Bezes, 2009). Cet effort n’avait certes que peu d’échos dans le débat public large. L’évaluation avait pourtant acquis droit de cité dans les pratiques éducatives et les tests sur les élèves se généralisaient dans l’esprit d’évaluer l’efficacité de l’enseignement primaire et secondaire.

En 2014, l’espace public est tout différent en France. C’est l’évaluation *dans l’Université* qui a acquis une grande notoriété et provoqué des polémiques très vives, au point que la notion même d’évaluation a été interrogée par un grand nombre de disciplines et de chercheurs, faisant l’objet de critiques radicales : elle est ainsi considérée par une part significative des universitaires français, comme « une des grandes impostures de la dernière décennie » (Zarka, 2009). Chez un grand public averti, dans le même temps, plusieurs économistes se sont fait connaître en France en proposant de révolutionner les interventions contre la pauvreté dans les pays très pauvres ou les politiques d’éducation en France. Dans des articles de la grande presse³, *Le Monde*, *Le Figaro* ou *Libération*, ils ont vanté les mérites d’une méthode qu’ils ont proposé d’étendre : l’expérimentation, parce qu’elle « marchait », par exemple pour soulager la pauvreté des Indiens. Ici il s’agissait plus précisément d’évaluation des politiques publiques et de réforme de ces politiques. On ne peut rêver plus grand contraste entre l’évaluation « imposture » des universitaires et l’évaluation expérimentale bienveillante des économistes de la pauvreté ou de l’éducation. Ajoutons

¹ Les données empiriques qui fondent l’analyse qu’on va lire ont été collectées, de façon participative, par l’auteur sur une période qui va de la fin des années 1980 à aujourd’hui. En 1988 l’auteur était responsable de la recherche à la Caisse nationale des allocations familiales et contribua à l’évaluation du RMI, en tant qu’administrateur. Dans divers organismes, il a commandité et réalisé de nombreuses études d’évaluation, dont la pratique l’amena parmi les fondateurs de la Société française de l’évaluation en 1999. Depuis cette époque, il a animé, puis co-animé des groupes de travail au sein de la SFE, dont le groupe intitulé « standards de qualité et déontologie ». L’analyse présente tire parti des centaines d’entretiens réalisés au cours de cette période et en partie synthétisés pour l’écriture de Barbier et Matyjasik (2010) et Barbier (2010).

² Pour emprunter à B. Jobert (1998) le concept que nous avons utilisé dans le domaine social, (Barbier, 2008).

³ Deux exemples typiques d’articles : Esther Duflo, interview dans *Libération*, « Je voudrais changer la perspective même de l’aide aux plus pauvres », le 6.1.2012 ; Marc Gurgand, *Le Monde* 14.10.2009, « Contre l’absentéisme scolaire, expérimentons d’abord ».

que le paysage des années 2010 au plan transnational est bien différent de celui des années 1990, du point de vue de l'évaluation : dans les forums spécialisés, le recours à ce qu'on a appelé « l'évaluation d'impact⁴ », se répand, une technique qui plaît aussi bien à la Banque mondiale qu'à la Commission européenne, ou au gouvernement mexicain par exemple.

Cette actualité bien différente de celle des années 1990 justifie qu'on revienne ici sur les « sources intellectuelles » de l'évaluation des politiques publiques (EPP, désormais en abrégé). On ne peut échapper pour cela à la question de la définition. La pratique de spécialistes qui s'appelle EPP est difficile à saisir rigoureusement, particulièrement en France, où le regard comparatif a tendu à montrer que l'EPP y était peu « institutionnalisée » (Jacob, 2005). Mais cette difficulté d'identifier une pratique n'est pas propre à la France. C'est ce que nous verrons dans une première partie.

Ensuite, nous posons la question des sources de l'EPP de deux façons. La première consiste à repérer, dans l'histoire des connaissances, de grands courants qui ont permis la construction, puis le développement de l'évaluation des politiques publiques. S'agissant de trouver des sources intellectuelles pour une pratique difficile à définir, on ne pourra manquer d'observer qu'elles sont multiples, internationales et partiellement contradictoires ; elles composent un ensemble d'influences disciplinaires variées, se formant sur une très longue période, qu'on peut faire remonter aux pratiques des sciences camérales en Allemagne au XVIII^e siècle. Elles constituent un méli-mélo pragmatique de références, variable selon les lieux, les cultures politiques et les époques. Dans un second temps, on se demande quelles sont les sources auxquelles se réfèrent effectivement les praticiens qui se réclament de l'EPP, mais pas seulement en France, grâce à une enquête dans plusieurs pays (Barbier et Hawkins, 2012) : l'observateur est frappé par la présence simultanée d'un « état de l'art » international, d'une part, et de pratiques variées de l'autre au plan national. Le repérage de ces deux types de sources, identifiées dans l'histoire des pratiques de l'EPP d'une part, puis, aujourd'hui, chez les praticiens, nous permet de mieux anticiper les chemins qu'en tant que « discipline », l'EPP est susceptible de prendre dans l'avenir, au croisement de l'utilisation des sciences sociales et de leur application spécialisée dans les politiques publiques.

1. Évaluation des politiques publiques : une définition impossible en sociologie ?

Pour situer la difficulté qui s'attache à la caractérisation de l'EPP sur la base de critères objectifs et indépendants de choix normatifs singuliers, on peut d'abord observer que l'EPP, à l'image des sciences de la gestion (du management), est soumise à la mode (Barbier, 1999). Là n'est pas l'essentiel cependant : l'essentiel tient dans le fait que l'EPP est *une démarche pratique, en soi non scientifique*, conduite par des consultants et des experts pour le compte de commanditaires qui sont des acteurs publics, parfois directement des hommes ou femmes politiques. Cette position de « croisement » entre la mobilisation des connaissances des sciences sociales et la réponse obligatoire à une commande de la part d'acteurs en position de décision publique accueille une large palette d'activités qui ont des buts similaires : elles visent toutes à rassembler et mettre en forme des *informations à caractère évaluatif* – c'est-à-dire des connaissances construites à propos des conséquences (les résultats, les effets) des interventions que les acteurs en position de décision ont conduites. Le décideur cherche

⁴ Voir la discussion de la notion plus loin.

dans tous les cas à comprendre les effets qui ont suivi la mise en œuvre de ses décisions. Plusieurs types d'activité sont susceptibles de répondre à ce besoin des commanditaires et bien des activités correspondent à la définition générique : par exemple, l'audit, le contrôle de gestion, la comptabilité publique, le contrôle de conformité, la recherche opérationnelle, y compris les sciences camérales allemandes (Laborier *et al.*, 2011). Parce que l'évaluation est, par nature, un champ éclectique, il est donc nécessaire de parler en même temps d'elle *stricto sensu* et d'autres activités voisines. *Stricto sensu*, l'évaluation d'une politique est en général définie par ses praticiens par son objectif, à savoir « apprécier l'efficacité de cette politique en comparant ses résultats aux objectifs assignés et aux moyens mis en œuvre » (Perret, 2001, p. 3). Par ailleurs, la division du travail entre ces activités voisines n'est pas homologue partout dans le monde⁵, et les controverses abondent pour établir des frontières. Les controverses elles-mêmes ne sont pas équivalentes dans les différents pays pour une raison simple, sur laquelle nous reviendrons en troisième partie, à savoir que l'EPP est construite, dans chaque pays, au sein d'une culture (ou de cultures) politique singulière.

Bien sûr, la solution existe toujours d'accepter, *a minima*, que l'évaluation est ce que font ceux qui disent qu'ils font de l'évaluation. Dans le monde des pays riches, surtout, mais de façon croissante dans les pays émergents, des communautés d'experts se reconnaissent en effet comme évaluateurs et, comme nous le verrons en troisième partie, ils font référence à un « état de l'art »⁶. Mais cette position n'est pas pleinement satisfaisante. Il faut considérer l'EPP comme une « profession » possédant ses spécialistes, son domaine de publication – pour l'essentiel en anglais – ses filières de formation, ses codes déontologiques, ses associations professionnelles, etc.. L'EPP est marquée par des « générations » d'approches, qui constituent autant de candidatures posées par leurs auteurs, à un moment donné, à la présentation d'une synthèse et d'un dépassement original de la littérature précédente⁷. Cette littérature vient nourrir le stock sans cesse renouvelé de « l'état de l'art ».

C'est pourquoi la concurrence et le renouvellement des méthodes expliquent par exemple, après bien d'autres péripéties depuis les années 1960, que « l'évaluation d'impact » soit devenue à la mode dans les années 2000. Il faut préciser les choses car ce vocable regroupe de nombreuses techniques. Disons qu'une version de l'évaluation recourant à la comparaison entre échantillons tirés de façon aléatoire (*random control trials*) devint célèbre en France dans les milieux de l'administration et du conseil. Au même moment, bien au-delà, en Amérique Latine par exemple, elle connaissait aussi un intérêt lié à la mise en œuvre des programmes dits de « *conditional cash transfers* ». Le succès de l'évaluation d'impact (au sens des *randoms controls trials*) était également porté par d'autres causes sociologiques, à commencer par l'évolution des financeurs de l'aide au développement qui critiquaient les limites des méthodes d'évaluation utilisées jusqu'alors dans ce domaine de financement crucial du point de vue international (Labrousse, 2010 ; Barbier, 2012).

⁵ Depuis de nombreuses années, par exemple, aussi bien aux États-Unis qu'aux Pays-Bas, les organismes qui contrôlent la dépense publique s'occupent à la fois de contrôle, d'audit et d'évaluation.

⁶ Un cas intéressant est le groupe international Inteval, fondé par Ray Rist (sociologue, longtemps spécialiste de l'évaluation à la Banque Mondiale) dans les années 1980 ; le groupe a publié collectivement 22 livres depuis les années 1990 aux éditions Transaction Publishers. En tant que membre du groupe, nous avons enquêté sur sa pratique depuis 1995 (enquête participante).

⁷ Un exemple parmi d'autres : Guba et Lincoln (1989), qui parlent de l'évaluation de « quatrième génération ».

Peu des utilisateurs de ce type « d'évaluation d'impact » savaient pourtant qu'une célébrité analogue autour de la « méthode expérimentale » s'était construite aux États-Unis, dans les années 1960. La « méthode expérimentale » de l'époque reposait précisément elle aussi sur le principe qu'on comparait des individus soumis à une intervention avec d'autres, strictement comparables toutes choses égales par ailleurs, qu'on maintenait à l'écart de l'intervention en question (échantillon de contrôle). C'est à cette époque que le psychologue Donald T. Campbell écrivit dans la revue *American Psychologist* qu'il était temps d'adopter ce qu'il appela « *a truly experimental approach to social reform* ». Pour lui, les politiques publiques devaient désormais faire l'objet de tests, et seules les expérimentations avec comparaison d'échantillons aléatoires étaient valides pour déterminer si elles pouvaient être efficaces : « *true experiments should almost always be preferred to quasi-experiments where both are available. Only occasionally are the threats to external validity so much greater for the true experiment that one would prefer a quasi-experiment* » (1969, p. 426). Il n'écartait donc pas complètement la « quasi-expérimentation », mais le reste des méthodes n'était pour lui que tromperie : « *a morass of obfuscation and self-deception* » (*ibid.*). Le propos de Campbell distingue les expérimentations aléatoires et les méthodes qui ne recourent pas au tirage au sort, mais qui – on les appelle quasi-expérimentales – utilisent un échantillon pour la comparaison qui est reconstitué après l'intervention, à partir de données existantes et où les individus sont strictement comparables comme dans le tirage au sort de la méthode expérimentale. Au-delà, une variété de méthode existe, certaines reposant sur des modèles économétriques, d'autres sur la comparaison avec contrefactuel, ou encore les méthodes des « doubles différences » ; un groupe de travail de la SFE a aussi, à juste titre nous semble-t-il, fait observer qu'on peut reconstituer de plusieurs manières une situation « contrefactuelle » pour faire la comparaison, comme c'est le cas dans la méthode « basée sur la théorie » (*theory-driven*) et l'analyse de contribution (SFE, 2011, p. 33-37). Plus généralement, ce groupe d'experts a fait remarquer que la question de « l'impact » est à la racine de la plupart des pratiques de l'évaluation.

Les utilisateurs d'évaluations et les évaluateurs connaissant cette variété de fait de l'évaluation d'impact ont donc été surpris quand, en France, les économistes de l'École d'économie de Paris (EEP) et du CREST (INSEE) ont, en 2007-2008, affirmé la supériorité absolue d'une méthode qu'ils tendaient à considérer comme la seule possible pour évaluer des impacts. Cette forme d'évaluation (il s'agissait d'échantillons tirés au sort parmi les demandeurs d'emploi) – déclara Marc Crépon (CREST INSEE) le 8 juin 2008, en présentant les résultats d'une très lourde évaluation des services pour l'emploi – « produit des résultats de façon transparente qui ont la force de l'évidence » et ces résultats alimentent les débats sur les politiques, mais ils ne « font pas débat ». Dans la même veine, son collègue Marc Gurgand écrivit⁸ : « L'évaluation ne se substitue pas au débat public. Mais elle lui apporte des faits, qui ne peuvent être construits qu'à travers l'expérimentation. Refuser de construire convenablement ces faits, c'est, d'une certaine façon, refuser le débat (...) Le scandale serait d'importer des dispositifs dont les effets potentiels ne sont pas connus, sans les expérimenter et sans les évaluer ». Selon ces auteurs, il n'y avait donc qu'une méthode, et c'est la meilleure (un « gold standard⁹ »), l'expérimentation. C'était

⁸ *Le Monde*, op.cit.

⁹ François Bourguignon, ex-économiste en chef de la Banque mondiale, intervient en mai 2008, dans un colloque organisé par le ministère du travail : il cite « la méthode expérimentale sur échantillons aléatoires » pour affirmer qu'elle « a été à juste titre qualifiée d'« étalon-or » de ces méthodes ». À l'époque il est directeur de l'École d'Économie de Paris.

aussi l'avis d'Esther Duflo, professeur à MIT, qui écrivait à la même époque : « les données non expérimentales, à quelques exceptions près, ne fournissent pas de résultats suffisamment rigoureux pour élaborer des politiques, ce qui constitue une motivation importante à mener des expérimentations » (Duflo et Banerjee, 2009, p. 694). Cette position promouvait donc une seule méthode (présentée comme la meilleure) et équivalait à écarter une bonne part de quarante ans de pratique de l'EPP dans le monde développé. Les tenants de la modélisation ne se privèrent donc pas de porter la contradiction rapidement (Heckman, 2010), on y fait allusion plus loin.

Au-delà de cet exemple illustrant les effets de mode au début des années 2000, la difficulté de rassembler sous une seule bannière de nombreuses variantes se réclamant de l'EPP confirme une observation sociologique de longue durée dans les milieux de l'évaluation en France et à l'étranger¹⁰ (Barbier, 2010 ; 2012 ; Barbier et Matyasik, 2010, Barbier et Hawkins, 2012) : une définition strictement sociologique de cette activité se révèle impossible. Au mieux, on peut souligner les traits distinctifs d'une large *famille d'activités* et en situer historiquement l'apparition et les développements. L'évaluation aurait été « inventée », dit-on, aux États-Unis, sous les présidences Johnson et Kennedy, dans les années 1960 : c'est la thèse qu'adoptent la plupart des travaux en langue française (Monnier, 1992 ; Perret, 2001). Il est sans doute plus exact de dire que l'EPP, telle qu'elle se pratique aujourd'hui en France, trouve une de ses sources dans l'exemple américain et qu'on en peut retracer la filiation avec les *policy sciences* dans ce pays (Spencehauer, 1995). L'évaluation constitue un éventail de pratiques des autorités publiques permettant de juger de la valeur (selon des critères multiples) des programmes, et, bien sûr, d'en tirer des conséquences opératoires. Il est possible de distinguer *a minima* quatre grands types d'activités d'où l'action publique tire des « informations évaluatives ». Le premier est la recherche scientifique appliquée (évaluée par les pairs dans chaque discipline universitaire). Le second est l'étude/recherche finalisée, qu'on pourrait dire « caméraliste »¹¹, à destination d'un ou plusieurs financeurs intervenant plus ou moins dans la commande et dans l'étude elle-même. Ensuite, il existe au moins deux catégories d'évaluations : des opérations, souvent multidisciplinaires, qui embrassent des politiques assez larges, sous l'égide d'instances *relativement* autonomes, plus ou moins formalisées, comme les commissions d'évaluation mises en place par l'ex-Conseil scientifique de l'évaluation (CSE) ou le Commissariat du Plan ; et, enfin des évaluations ou activités voisines (pour l'essentiel économiques et sociologiques) à destination d'un commanditaire unique (on retrouve ici l'audit et le contrôle de gestion déjà mentionnés).

Cette typologie, simplificatrice, sépare en tout cas clairement l'EPP de la recherche et des études. Cela fait en général consensus : la recherche est à part, car elle pose ses propres questions et l'étude, de son côté n'est pas vraiment aimantée par un questionnement spécifique à propos *des effets* de l'action publique. Mais, pour le reste, les avis sont variés et les définitions de l'évaluation, relatives : on peut arguer, certes, de l'importance, dans l'EPP de la question de la causalité (Barbier, 2010). Mais l'ambiguïté des définitions est aussi historique : il est possible – à la limite – de tenir une position comme celle de V. Spencehauer, pour qui l'évaluation est une question aussi vieille que celle de l'État (1995, p. 24). Aussi bien Colbert qui, en France réfléchit sur la façon de favoriser l'augmentation de la population et les grandes familles, que Goethe, en Allemagne, conseillant le duc de

¹⁰ Voir note 1.

¹¹ Pour reprendre la formule de R. Boudon (2003).

Weimar, peuvent être alors considérés parmi les ancêtres de l'évaluation. Enfin, la variété des activités voisines de l'évaluation, traitant « d'information évaluative » s'articule avec une donnée géographique et culturelle à la fois : pour chaque situation nationale donnée, il existe une insertion spécifique des pratiques d'évaluation dans une culture politique particulière (Barbier et Hawkins, 2012).

2. Ingénierie sociale et management public

Après le constat sociologique de la fondamentale variété d'application du terme évaluation des politiques publiques, venons-en maintenant à la recherche de ses sources intellectuelles. Il faut alors distinguer, ne fût-ce qu'analytiquement, entre les pratiques et l'activité de réflexion qui les accompagne. En tant qu'*activité de réflexion*, il nous paraît possible, sans en méconnaître les spécificités, de rattacher pleinement l'EPP à l'ensemble composite des « sciences de gestion », comme on les désigne en français¹².

En effet, comme la gestion (Barbier, 1999), et la gestion publique en particulier, l'EPP relie des connaissances scientifiques, des pratiques, et des discours normatifs à propos de ces dernières. En EPP, les connaissances font, comme en gestion, l'objet de construction d'outils. De la même façon qu'il y a des connaissances scientifiques (sciences de gestion) à côté de pratiques (la gestion publique et la gestion privée), l'EPP est faite de réflexion d'un côté et de conduite d'opérations de l'autre. En France, depuis les années 70-80, les acteurs sociaux pratiquent ce qu'ils considèrent couramment comme « gestion publique »¹³. Par « management public », on désigne ici une activité composite¹⁴, à la longue et complexe histoire, chevauchant plusieurs traditions nationales. Cette histoire est d'ailleurs souvent conçue très différemment selon qu'on prend le point de vue des sciences de gestion (du management¹⁵) ou celui de la science et la sociologie politiques.¹⁶ Nous nous centrons ici sur l'EPP comme l'une des activités parmi le groupe de celles qui sont liées à la mise en œuvre, au pilotage et à l'appréciation des politiques, considérée du point de vue privilégié des gestionnaires qui doivent rendre compte de leur gestion publique. Si l'on suit l'observa-

¹² On peut noter que les différentes traditions nationales ne font pas des classifications homogènes de ces « sciences », sans pouvoir approfondir la question.

¹³ Jean-Pierre Nioche (1982, p. 52) parle pour l'évaluation de « nouvel outil de la gestion publique ».

¹⁴ C'est-à-dire un ensemble d'activités incluant les pratiques de gestion et d'administration dans le secteur public, depuis les réflexions américaines pionnières des années 1950 jusques et y inclus le « new public management » contemporain. Cette conception s'oppose par exemple à l'usage restrictif d'un Philippe Bezes (2009) dans son panorama de l'administration française des années 1960 aux années 2000 : pour lui, le management public est l'un parmi quatre réseaux d'expertises, dont l'évaluation et la RCB (rationalisation des choix budgétaires), les cercles de qualité, et la sociologie croziérienne. Il tend d'ailleurs à identifier cette acception du management public plutôt au « contrôle de gestion » (ibid., p. 290). Par contraste, Patrick Gibert (2008, p. 18) reprend à son compte une définition *mainstream* de l'*Oxford Handbook of Public Management*, comme un domaine à la fois universitaire et « de pensée et d'activité d'application pratique ».

¹⁵ R. Laufer a noté que l'usage de la notion de « management », dans le sens du management public n'est pas aussi ancienne qu'on le penserait et date, aux États Unis, des années 1950 (2008, p. 34, s'appuyant sur Van Riper, 1983). Il a abondamment montré que la distinction entre public et privé est fondamentalement relative en matière de sciences et pratiques de gestion.

¹⁶ Sur la notion de management public, *lato sensu*, voir le numéro de *Politiques et Management Public* 26 (3) de 2008.

tion essentielle de Romain Laufer quant à l'expansion contemporaine de la catégorie de ce qui est « public », et quant aux problèmes de légitimité désormais généralisés au public et au privé, l'EPP devrait, en théorie au moins, se trouver à la pointe du management public, puisqu'elle se préoccupe au premier chef de la façon de rendre des comptes (*accountability*).

Peut-on capturer, de façon synthétique, l'inspiration d'une pratique aux frontières floues ? Nous nous concentrons, pour simplifier une histoire complexe qui traverse plusieurs disciplines, sur deux sources essentielles : la première peut être décrite comme une ingénierie sociale (*social engineering*), bien représentée, dans les années 1960, par l'interprétation scientifique du psychologue D. T. Campbell. La seconde ressortit au vaste ensemble du management public tel qu'il s'invente aux États-Unis dans les années d'après guerre, évoluant de façon incessante, en liaison plus ou moins étroite avec l'analyse des politiques publiques (la « *policy analysis* »), en diversifiant ses champs d'application ; plus ou moins clairement identifié, un corps de doctrine renouvelé y prend naissance plus tard, dans les années 1990, sous l'étiquette « *New Public Management* » – NPM.

2.1. Une ingénierie sociale éloignée de la causalité complexe de la politique

Laissant de côté l'histoire plus ancienne des activités qui précèdent la dénomination *contemporaine* d'EPP, on peut rappeler que la littérature d'évaluation fait en général remonter la filiation de la « discipline » aux premières applications des sciences sociales au XIX^e siècle¹⁷, à la naissance du calcul économique, avant même que soient formés la notion et le concept même de « politique publique ». Cette époque est celle où un certain scientisme laisse encore espérer que l'humanité va prendre progressivement la maîtrise des activités politiques, grâce à la science, selon un progrès continu. Le saint-simonisme et le comtisme fournissent aux sciences sociales cette dimension qui, si elle a désormais été discréditée, n'en reste pas moins présente à l'état latent : au sein de la « discipline » évaluation, cette tentation est présente depuis les origines¹⁸. Carol H. Weiss (1998, p. 11-13) a noté que, dès les premières années du XX^e siècle, des études sociales rudimentaires furent menées dans les domaines de la santé et de l'éducation afin d'augmenter l'efficacité des interventions publiques. De premières expérimentations psychologiques avec groupe de contrôle aléatoire eurent lieu dans les années 1890 en Grande-Bretagne. Dans les années 1920 et 1930, on teste la taille des classes, la supervision des écoles, l'éducation à la santé en milieu scolaire, etc. (Monnier, 1992, p. 41). L'idée sous-jacente est que si l'on manipule adéquatement ces variables par des interventions, on sera à même de produire *des résultats sociaux* ; c'est le sens même de l'ingénierie sociale dont l'histoire est, elle aussi, multiple. Jusque dans les années 1970, dans les pays démocratiques (et bien sûr dans les pays soumis à la dictature) des expérimentations sociales sont d'ailleurs menées, parfois avec des cobayes humains, parce qu'une croyance existe dans ces sociétés selon laquelle l'intervention politique peut être organisée comme on le fait pour mettre en œuvre des machines, d'où le nom d'*ingénierie*

¹⁷ Dans l'aire française, on a évoqué, parmi les ancêtres de l'évaluation, des gens comme Quételet et Villermé.

¹⁸ Cette dimension est tout aussi présente dans l'usage du comportementalisme (*behaviourisme*) dans de nombreux travaux économiques contemporains. Un exemple typique est l'ouvrage à fort succès de S. D. Levitt et S. J. Dubner, 2006, *Freakonomics*, Paris, Denoël, Folio.

(King, 1999)¹⁹. Le terme *ingénierie* est d'ailleurs resté aujourd'hui, un terme courant du management²⁰. Le terme « social engineering » prendra progressivement une connotation péjorative, faisant référence à des expérimentations éthiquement condamnables²¹, comme les expériences d'eugénisme (King, 1999), mais aussi les « réformes de la pensée » dans les pays communistes, en Chine en particulier (Barbier, 1982).

Même si ses partisans ont défendu l'idée que D.T. Campbell a été accusé à tort de scientisme, il reste dans les années 1960 un inspirateur essentiel d'une conception scientiste de l'EPP. Il conçoit « *the experimenting society* » (Campbell, 1969)²² sur la base de nombreux postulats scientistes : la société deviendrait une société « scientifique » avec les valeurs scientifiques d'honnêteté, d'ouverture à la critique, d'absence de dogmatisme... une société qui s'engagerait à « éviter les illusions ». Vincent Spence a noté que Campbell a joué un rôle clé dans l'évaluation américaine car « *les premiers chercheurs-évaluateurs américains ont commencé par techniciser la program evaluation selon le modèle très scientiste et béhavioriste de l'expérimentation sociale élaboré dans le champ de la psychosociologie* » (1998, p. 91-92). Dans leur manuel de référence, Peter H. Rossi et Howard H. Freeman décrivent, avec le recul des années 1990, le personnage : « *Selon lui, les politiques devraient sortir de l'expérimentation constante de moyens pour améliorer les conditions sociales* » (1993, p. 29). C'est le monde entier qui devient alors un laboratoire : « *the community and the nation, if not the world, should be seen as a laboratory for social experiments* » (*ibid.*). C'est dans cette orientation que les expérimentations contrôlées aléatoires se sont développées aux États-Unis dans les années 1960, au moment même où naissait et s'installait solidement l'EPP dans le cadre de la *Great Society* de l'administration Johnson, mettant en œuvre des programmes sociaux très ambitieux.

Un point est ici important pour la comparaison avec les développements actuels de l'évaluation par expérimentation aléatoire contrôlée : la question de la causalité. L'expérimentation issue des méthodes campbelliennes, dans les années 1960, se fonde sur une conception de la causalité mise en évidence, toutes choses égales par ailleurs, par la différence des caractéristiques mesurées (revenu, travail, santé, compétences...) entre un groupe « de traitement » et un « groupe de contrôle »²³. Or, cette conception de la causalité est désormais critiquée aussi bien en économie qu'en sociologie. Par les économistes, d'un côté, comme

¹⁹ Desmond King (1999) a concentré son enquête sur les programmes « contraires au libéralisme » (« illiberal ») de la Grande-Bretagne et des États-Unis ; il analyse ainsi, dans l'Angleterre et le Pays de Galles, une expérience typique de cette « ingénierie » pour la stérilisation de personnes considérées comme « faibles d'esprit » ; de même il décrit, à la même époque des années 1930, la mise en place de « camps de travail » pour « réduire » les chômeurs aussi en Grande-Bretagne. Des expériences américaines sont comparées. L'ouvrage n'aborde pas, mais ils existent d'autres types de programmes d'ingénierie sociale dans les pays scandinaves ou aux Pays-Bas. En revanche il rattache le « workfare » contemporain à cette tradition.

²⁰ A. Hatchuel (2008, p. 10) souligne que les sciences de gestion se pensent même dans les années 50-60, comme « une 'ingénierie' qui tire sa légitimité de doctrines reconnues et de pratiques validées en entreprise ».

²¹ D'où la connotation péjorative du terme anglais « *social experimentation* », que n'a pas le français « expérimentation sociale ». La Commission européenne, dans son jargon, emploie pourtant le terme anglais pour son programme dit Plateforme pauvreté, dans le cadre de sa « stratégie » Europe 2020.

²² Il donnera son nom à une organisation aujourd'hui très puissante qui s'occupe de rassembler et d'analyser les études d'évaluation dans le domaine du crime, de l'éducation et des affaires sociales ; l'organisation dite « *Campbell Collaboration* » a été créée en 2000.

²³ Cette façon de procéder a, historiquement, pris une place centrale dans l'expérimentation médicale (Labrousse, 2010).

Heckman (2010, p. 1-3) pour qui l'objectivation d'une causalité suppose l'application d'une modélisation pratiquée sur les données, ce que ne permet pas le traitement « statistique », sans théorie particulière, qui est le propre de l'échantillonnage aléatoire. Mais, plus largement, par la tradition weberienne qui récuse une causalité mécanique simple. M. Weber (1996, p. 56-57), par exemple dans *L'Éthique protestante et l'esprit du capitalisme*, tient que la causalité, intrinsèquement complexe, apparaît comme un « *réseau de connexions causales entremêlées à l'infini* ». Cela ne le conduit nullement à abandonner la catégorie centrale de la causalité, mais plutôt à en spécifier des modalités d'application qui diffèrent de celles des sciences exactes. Contrairement aux raisonnements univoques, le chercheur ne trouve jamais au fond une seule réponse causale, mais des « séries causales » qui visent un « *regressus causal de plus en plus large, sans qu'il puisse jamais être mené jusqu'à son terme* », comme l'explique Jean-Pierre Grossein dans sa préface à l'ouvrage cité.

2.2. Management public, policy analysis, vers une vision plus réaliste et pluraliste

L'EPP n'en est pas pour autant restée à l'enseignement campbellien de l'expérimentation et son inspiration s'est diversifiée considérablement, ce qui a amené son succès indéniable à travers le monde développé. Toutefois, la croyance d'un progrès en quelque sorte inexorable de la gestion publique et d'une maîtrise par la bureaucratie rationnelle restera longtemps présente aux États-Unis. Ainsi que l'a noté, au début des années 1980, P. Van Riper, dans une analyse rétrospective de l'évolution de l'administration américaine depuis le XIX^e siècle, l'idée d'un progrès continu de l'administration moderne y était implantée depuis la fin de la première guerre mondiale. Et le management public (« management analysis ») se développa dans l'idée que « *the modern administrative state, properly directed, could accomplish almost anything we might envision (...) by 1945, for the first time in our history, the positive administrative state was perceived as a positive good by a large proportion of the voting public* » (1983, p. 485). Nous n'avons pas la place de rentrer ici dans les détails de cette histoire complexe (Spencehauer, 1999, voir aussi Thoenig, 2008) qui mêle travaux d'analyse des politiques et mise en place de programmes comme le PPBS (Program, Planning, Budgeting System, finalement supprimé en 1972, et qui inspirera le programme RCB en France).

En tous les cas, dès la fin des années 1970, les grands programmes mis en œuvre par l'administration se révèlent décevants par leurs résultats, donnant lieu aux réflexions pionnières sur la mise en œuvre (Pressman et Wildavsky, 1979). Ainsi, si une combinaison « rationnelle » de management public et d'expérimentalisme a pu, un temps, soutenir la réflexion et la pratique américaine de gestion des politiques publiques, cette époque est révolue, notamment parce que les travaux d'analyse des politiques publiques en ont remis en cause le simplisme (Nioche, 1982 ; Palumbo, 1987). Par exemple, au début des années 1990, les auteurs d'un célèbre manuel d'évaluation des politiques, P. H. Rossi et H. H. Freeman (1993, p. 29), opposent une version « pragmatique », donc réaliste, des potentialités de l'évaluation et passent en revue la palette des nombreuses méthodes disponibles. De leur côté, les spécialistes de l'évaluation des années 1980, comme C. H. Weiss, montrent des attentes modestes quant aux potentialités de l'EPP (et, plus généralement, de l'application des sciences sociales aux politiques publiques). En France, l'échec de la RCB, importée

des États-Unis, fait aussi partie de ce désenchantement, qui rebondira cependant dans une relance de l'EPP (Spencehauer, 1999). Finalement, dans l'EPP, aux États-Unis et ailleurs où elle s'est développée, un double constat était fait dès les années 1980 et, *a fortiori* 1990, de façon consensuelle : les illusions de l'expérimentalisme étaient enterrées, car cette méthode ne pouvait s'appliquer que dans des conditions bien précises et limitées ; de même les illusions des sciences de l'ingénieur apparaissaient clairement quant aux possibilités d'optimisation des politiques, face aux sciences de l'organisation, à la sociologie et à l'analyse des politiques. Les méthodes se diversifiaient et les débats étaient vifs dans la communauté internationale restreinte d'experts se réclamant de l'EPP : prise en considération des utilisateurs, évaluation basée sur des théories d'action des politiques (normative et descriptive), construction de contrefactuels, réflexion sur les méthodes d'attribution des effets précis à des causes, synthèse réaliste de plusieurs méthodes fondée sur l'identification des mécanismes (Pawson et Tilley, 1997), et bien d'autres encore.

2.3. Le New Public Management et les politiques fondées sur les preuves (evidence-based)

Dans le domaine précis d'intersection entre évaluation et management public, ce sont sans doute les politiques « basées sur les preuves » (*evidence-based*) qui représentent le point le plus avancé contemporain. Un puissant mouvement social s'est en effet entiché, depuis les années 1990, de l'*evidence-based policy*, dans les pays anglophones et bien au-delà, qui associe « benchmarking » et économétrie (Desrosières, 2008, p. 28-32). Il s'est particulièrement développé en Grande Bretagne avec l'arrivée au pouvoir du *New Labour*²⁴. Les acquis en quelque sorte « réalistes » de la réflexion sociologique et de science politique précédente sur les politiques publiques ont semblé ainsi remis en cause. Ce mouvement international est d'une ampleur trop grande pour que nous l'analysions dans ses détails et ses déterminants multiples : il concerne aussi bien l'évaluation des politiques sud-américaines de « *conditional cash transfers* » ; les expériences lancées aux États Unis par les fondations philanthropiques (action du maire Bloomberg à New York) ; la promotion de « l'expérimentation sociale » par la Commission européenne au sein de son programme Europe 2020 ; la rediffusion des méthodes de tirage aléatoire d'échantillons en Europe, le lancement de programmes d'assistance sociale, comme le RSA en France, la réforme Hartz en Allemagne etc.. L'une des révolutions paradigmatiques les plus étonnantes dans ce domaine est représentée par la Communication de la Commission européenne Com (2013) 83 final du 20.2.2013 intitulée « *Towards Social Investment for Growth and Cohesion* ». Ce texte programmatique de réforme de la protection sociale prône la reconfiguration de cette dernière en programmes « ciblés, conditionnels et simples », mais aussi « de durée temporaire ». Ces programmes seraient ainsi plus facilement évaluables par des méthodes d'évaluation d'impact. Il faudra voir, dans le futur, comment une communication d'importance stratégique pour les politiques sociales en Europe contribue ou non à la modification de ce secteur en profondeur.

Force est en tous les cas de constater, que, dans l'EPP, on se trouve désormais bien loin de l'analyse des politiques publiques telle qu'elle était promue dans les années 1980 et 1990. Pourtant, comme dans le cas des limites de la « logique de l'ingénieur », les potentialités

²⁴ Pour le rôle du *New Labour* dans ce mouvement, voir parmi de nombreux articles : (Scolesbury, 2001).

de la conception de politiques fondées sur de véritables « preuves » sont à relativiser à la lumière du passé. C'est ce que fait, à l'aide de sa riche expérience de l'EPP, Carol H. Weiss en reprenant récemment le fil d'une vision *modeste et réaliste de la pluralité des méthodes*, après avoir depuis longtemps milité pour l'*apport cognitif* (Weiss, 1993), mais non immédiatement utilisable, que les opérations d'évaluation sont à même d'offrir à la société. Réfléchissant sur la mode adoptée par le ministère de l'éducation américain en 2003 de se fier uniquement à des programmes qui ont été évalués selon les méthodes expérimentales, elle pouvait rappeler, avec ses collègues, qu'à quelques exceptions près : *“many investigations in many countries have shown that evaluation has usually had only modest influence on policy and practice”* (Weiss et al., 2008, p. 30). L'enseignement fondamental des rapports entre l'évaluation et la politique n'était-il pas que *“evaluative evidence is not the only important component of policy making. The ends to which policy and programming are directed are at least as important as the means for getting there. Evidence-based policy assumes that the ends are given, and the issue is to choose the best way to get there. But there are circumstances where appropriate ends are much in contention and ethical and moral considerations must come into play”* (ibid., p. 44).

3. Les références de la pratique actuelle des évaluateurs en Europe et aux États-Unis

Pour compléter cet examen des sources intellectuelles de l'EPP, au moment où ses références semblent un peu partout se renouveler, il n'est pas inutile d'observer les variations nationales des pratiques d'évaluation aujourd'hui. C'est ce que nous permet une étude portant sur un grand nombre de pays (Barbier et Hawkins, 2012), étude à laquelle ont contribué certains des experts qui avaient permis, au début des années 2000, l'élaboration d'un « Atlas » de l'évaluation dans le monde (Furubo et al., 2002). Ses conclusions montrent que la pratique de l'EPP reste très diversifiée selon les États-nations. L'existence de cette variété est reconnue par les analystes du management public (Duran, 2010 ; Gibert, 2008 ; Hatchuel, 2008 ; Bezes, 2009). Les travaux empiriques centrés sur la pratique de l'EPP convainquent que, pas plus qu'il n'existe une gestion publique universelle ou uniforme, on ne peut espérer découvrir des références universelles pour l'EPP. Mais ce constat n'entre nullement en contradiction avec l'existence d'un « état-de-l'art » transnational de l'EPP : les deux aspects sont complémentaires.

3.1. Un état de l'art décontextualisé et des pratiques nationales diversifiées

L'inspiration – les sources intellectuelles – de l'évaluation se déploie dans un forum transnational permanent où sont débattues idées et théories de l'évaluation ; ce forum fonctionne comme un sous-secteur du forum international des idées et théories du management. Il mêle les caractéristiques des forums de communauté de politique publique et des forums scientifiques²⁵ que Jobert a identifiés (Jobert, 1998). Depuis la fin de la deuxième guerre mondiale, la production essentielle de recettes et théories s'y effectue en anglais et le rôle

²⁵ Le groupe Inteval cité plus haut est typique du fonctionnement de ce forum, réunissant des universitaires, des experts gouvernementaux et des consultants de plus de 20 pays.

d'acteurs internationaux (OCDE, Banque Mondiale, grandes fondations caritatives, etc..) y est essentiel. Le forum est peuplé d'un grand nombre de consultants, d'experts et de chercheurs en sciences sociales – l'économie y est probablement la discipline dominante, mais certains secteurs spécifiques sont également très actifs (ainsi, la criminologie, et les sciences de l'éducation). Les associations nationales ou internationales d'évaluateurs jouent aussi un rôle clé dans ce forum (ainsi l'European Evaluation Society, l'American Evaluation Association (AEA)), ainsi qu'un petit nombre de revues de référence (comme l'*American Journal of Evaluation*, ou *Evaluation*, pour n'en citer que deux). Ces références servent de medium de contact dans un vaste et diversifié système d'institutions de formation et d'unités universitaires, au sein desquelles un pays comme la France commence à peine à s'insérer (si l'on excepte un tout petit nombre d'universités). En constant renouvellement, le corpus de textes (articles, rapports²⁶, ouvrages) et d'outils ainsi rassemblés constitue, *dans les faits*, l'état de l'art international de l'évaluation, auquel les évaluateurs professionnels du monde entier sont censés se référer²⁷. Illustrons le point par deux exemples. Le premier est un article de la revue de l'AEA, le second un ouvrage. Dans l'article, F. Lleuw et ses collègues (van der Knap *et al.*, 2008) à l'aide d'une analyse de méta-évaluation, plaident pour une synthèse entre le recours aux critères de qualité de l'organisation Campbell Cooperation²⁸ et la méthodologie proposée par R. Pawson et N. Tilley, qui combine l'usage des méthodes expérimentales et des méthodes qualitatives. L'ouvrage en question (Pawson et Tilley, 1997) est typique d'une méthodologie sophistiquée produite par les deux auteurs sur la base d'une approche sociologique, d'un bilan approfondi de pratiques, ainsi que d'une critique systématique des fondements épistémologiques et méthodologiques de l'évaluation : cette approche a pour principe de mettre au centre du questionnement évaluatif les « mécanismes » complexes par l'intermédiaire desquels le programme peut avoir des conséquences. Cette question est également travaillée par d'autres auteurs qui distinguent la causalité « simple » des causalités « complexes » et « compliquées » (Rogers, 2008).

C'est sur la base de tels travaux, qui constituent un corpus de recettes, d'outils et d'ouvrages que se construit et se diffuse un « état de l'art » de l'EPP, sans cesse renouvelé, mais dont la capacité d'innovation technique et, surtout, la capacité de cumulation de connaissances n'est pas toujours aisée à identifier : l'essentiel du travail de recherche dans l'EPP, en effet, s'accomplit au sein des disciplines classiques (sciences de gestion, économie, sociologie), pour ensuite être appliqué dans la conduite d'opérations d'évaluation. Ce corpus est mis en œuvre par des experts, travaillant ou se formant en anglais, enseignant dans le monde entier²⁹. Pendant ce temps, les

²⁶ Un exemple de rapport influent est celui publié par la Banque mondiale et la fondation B. et M. Gates et la fondation Hewlett Packard, en 2006, « *When will we ever learn? Improving lives through impact evaluation* » (Center for Global Development, Report of the Evaluation Gap Working Group, Washington).

²⁷ C'est un article standard des codes de déontologie de l'évaluation que de prescrire pour les évaluateurs une connaissance approfondie de l'état de l'art international (voir, pour un exemple, le principe de compétence de la Charte de l'évaluation des politiques et des programmes publics de la Société française de l'évaluation (SFE)).

²⁸ Déjà mentionnée, l'organisation, très influente, est sponsorisée par des fondations et des gouvernements et elle a pour mission l'établissement systématique de revues d'évaluation (méta-analyse) aux buts de fournir des standards de méthode, de qualité et des cumulations de résultats d'évaluation, dans le domaine du crime, de l'éducation et des politiques sociales.

²⁹ Un exemple est Ray Rist, ancien responsable de l'évaluation à la Banque mondiale, formé à la sociologie et spécialiste du management, consultant international de haut niveau, déjà cité.

pratiques nationales de l'EPP restent contrastées, voire disparates : l'idée qui a pu être un temps caressée (Furubo *et al.*, 2002 ; Toulemonde, 2000) d'une convergence des pratiques et systèmes d'évaluation en Europe ne s'est pour l'instant pas matérialisée. Le paradoxe apparent tient donc dans le fait que coexistent une couche transnationale, éditée en anglais international, référence décontextualisée de management public, et une pratique localement adaptée, qui est construite par la juxtaposition (et l'interpénétration) des références nationales. On prendra, pour terminer cet article quelques illustrations de cette diversité des « systèmes institutionnels de l'évaluation » dont les traits variables s'expliquent par leur insertion dans des cultures politiques et des institutions elles-mêmes résistantes à l'homogénéisation, toujours possible en théorie, mais jamais réalisée en pratique, par l'influence des références internationales.

3.2. Les cas de la France, de l'Italie et de l'Irlande : illustrations

La pratique de l'EPP qui s'est installée en France à partir des années 1980 (Nioche, 1982 ; Barbier, 2010 ; Barbier et Matyjasik, 2010) est née avec de grandes difficultés d'institutionnalisation, en raison du rôle de l'État et des grands corps d'expertise de la haute fonction publique, globalement hostiles (Spenehauer, 1998). Une ouverture à l'EPP de la part de hauts fonctionnaires, censés être par principe des défenseurs de l'intérêt général, a pris beaucoup de temps à se mettre en place. Même sous la puissante incitation du développement des marchés de l'évaluation (avec le rôle de prescripteur de l'Union européenne) l'EPP en tant qu'*activité spécifique* n'a pas vraiment trouvé son autonomie. La commande qui résulte de la « modernisation » de l'État est autrement plus puissante (Bezes, 2009). Dans les années 2010, la décision de recourir à des opérations importantes reposant sur le tirage au sort et la comparaison des bénéficiaires et des non-bénéficiaires a représenté une réelle rupture de tendance, il est vrai non spécifique à la France. On a en effet observé un engouement nouveau, par exemple en Allemagne (à l'occasion de la réforme Hartz), mais aussi dans les pays scandinaves, pour ces méthodes. Enfin, en France où une référence à « l'évaluation pluraliste » est restée importante et constitue une spécificité en comparaison internationale, un débat jamais vraiment tranché oppose des tenants d'une évaluation « démocratique » à une évaluation « managériale », alors même que, dans les faits, comme on vient de le voir, l'EPP est, structurellement amenée à s'inscrire dans le management public. La situation est bien différente en Italie, où, comme l'explique N. Stame (2012), la pratique de l'EPP est étroitement liée à une culture juridique d'État et a été introduite lors de la mise en œuvre de réformes de l'administration restées relativement peu opérantes face à la logique de long terme de l'État italien. Ainsi, comme l'écrit l'auteur, les unités chargées dans l'administration de l'EPP sont souvent peu vivantes et routinières. Les formes du centralisme de l'État italien et de l'État français sont très différentes, le second étant beaucoup plus puissant relativement à la société civile que son équivalent transalpin dont l'histoire est plus courte. Le cas irlandais est tout à fait différent. R. Boyle et ses collègues (2012) montrent qu'une source politique du développement de l'EPP dans le pays a été liée à la mise en pratique de « partenariats » avec les acteurs sociaux, et, au tout premier plan, les partenaires sociaux, en particulier dans les politiques sociales et d'éducation, une pratique innovante et décentralisée. Cela n'entraîne pas, pour autant, l'existence d'une « culture irlandaise d'évaluation » unique. Mais, le contraste est fort, pour un pays de langue anglaise, avec la tradition du voisin britannique. Toutefois, ces auteurs montrent aussi qu'il ne saurait y avoir, en matière d'évaluation, de

déterminisme « culturel » de nature essentialiste. En effet, depuis les premiers signes de la crise économique de 2008, un puissant changement s'est manifesté en Irlande, adoptant des standards beaucoup plus centrés sur la nécessité de « preuves », du *New Public Management* et sur une centralisation accrue de l'EPP par l'administration. Dans ce dernier cas, comme dans le cas du lancement d'opérations avec tirage d'échantillons aléatoires en France, nous observons des interactions précises entre la couche décontextualisée de l'état de l'art, et les pratiques et institutions nationales. Ce fonctionnement est structurel et pourrait être illustré par de nombreux exemples pris dans de nombreux pays.

3.3. Le rôle des cultures politiques

L'*Atlas of Evaluation* (Furubo et al., 2002) l'avait déjà démontré : des relations étroites existent entre la façon dont l'EPP est conçue, sa formalisation, sa professionnalisation et les institutions et cultures politiques du pays concerné. Le point a été en particulier exploré dans une étude comparative entre la Suisse, la Belgique, les Pays Bas et la France (Jacob, 2005). La culture politique en cause n'est pas une culture essentialiste que l'on attribuerait à tel ou tel peuple, mais un système qui se construit dans une longue durée historique, en même temps que les institutions démocratiques, et qui est marqué par les empreintes du passé ; la culture politique relie des pratiques qui sont en quelque sorte « cristallisées » ou incorporées dans des institutions politiques et qui finissent par donner lieu à des valeurs et normes propres à une communauté donnée. C'est pour cela qu'à chaque fois, l'EPP s'insère dans un contexte culturel-politique qui la détermine et la contraint. La « couche » de l'état de l'art décontextualisé que nous avons identifiée précédemment se trouve en constante interaction avec le système national de références considéré. En outre, la tonalité singulière qui prévaut dans un pays ne signifie nullement qu'une culture nationale unique s'impose et s'oppose trait pour trait à une autre culture nationale. Nous avons ainsi montré que les secteurs de l'éducation, ou encore de l'aide au développement, voire des politiques de l'emploi, comportent, en matière d'EPP des traits qui traversent aisément les frontières nationales (Barbier et Hawkins, 2012). Les sociétés contemporaines qui pratiquent l'évaluation sont, en outre, proches les unes des autres par bien des traits et partagent souvent beaucoup de valeurs, si l'on reste à un grand niveau de généralité. Mais l'uniformité ne règne pas pour autant car la pratique de la politique est inscrite dans ces communautés qui passent leur temps à se différencier les unes des autres, souvent même à se différencier politiquement pour construire leur identité. L'EPP n'échappe pas à cette dynamique, qui chevauche à la fois les rivalités et les échanges, l'apprentissage et le cloisonnement entre nations. Les pouvoirs du parlement national, la place et l'autorité de l'administration, son degré de centralisation, ses modes de recrutement, l'histoire de l'État, etc., sont des traits pertinents pour expliquer les différences de systèmes et de cultures d'évaluation. Un autre trait illustre de façon claire la persistance de la diversité dans un monde qui, pourtant, possède un corps transnational de références : c'est celui de la diversité des standards de qualité et des codes de déontologie. Dans ce domaine, des efforts constants sont faits et sans cesse remis sur l'ouvrage, mais, malgré de puissantes références communes existantes³⁰, la diversité des chartes et des codes nationaux est une réalité avérée.

³⁰ On pense ici tout particulièrement à la liste de standards établis par le *Joint Committee of Standards for Educational Evaluation* américain.

Conclusion

En 2014, l'évaluation est pratiquée partout en Europe et aux États-Unis, mais elle n'est ni uniforme ni assurée dans un corpus véritablement *disciplinaire*. Si elle n'est pas, au sens français, à proprement parler une discipline, elle se nourrit, comme sa discipline d'accueil, le management public, à de nombreuses sources. Un parallèle pourrait aussi être tenté entre elle et une autre discipline éclectique dans un tout autre domaine, l'ergonomie (qui fait appel à la médecine, à la sociologie et la psychologie). D'où la difficulté, relatée ici, d'une définition stricte de l'EPP en sociologie, car elle se présente plutôt comme une palette d'activités visant à donner aux décideurs, et au-delà, aux citoyens, des informations construites pour apprécier les conséquences de leurs décisions. Pour autant, malgré des frontières floues, l'évaluation, cela ne fait pas doute, est une profession, qui a ses cursus de formation, ses recettes, ses organes de représentation, ses références déontologiques, et, bien sûr ses marchés. Cette profession est suffisamment libérale pour accueillir en son sein une grande variété d'experts qui partagent quand même une histoire et des sources d'inspiration.

L'expérimentalisme issu de la psychologie sociale en est une, essentielle, qui entretient une forme de nostalgie scientiste, parfois transformée en argument de vente à l'égard de d'hommes et femmes politiques cherchant à instrumentaliser les sciences sociales. Mais, la plus importante source d'inspiration de l'EPP relève de l'immense variété du management public et de l'analyse des politiques publiques. À la différence de l'expérimentalisme, ce corpus appelle au réalisme et à la modestie des demandes faites à l'EPP. C'est d'ailleurs une forme de réalisme, d'adaptation opératoire, qui explique qu'un solide corpus de référence développé par les experts et consultants en évaluation se trouve, à chaque fois, soumis au filtre exigeant des institutions et cultures politiques nationales, entretenant une diversité persistante.

S'agissant essentiellement d'une pratique orientée vers l'aide à la décision publique, l'EPP ne cesse de se transformer. Comme l'indiquent les titres des ouvrages de référence qui appartiennent à l'état de l'art que la profession internationale des évaluateurs reconnaît, l'histoire a produit des « générations » de références intellectuelles. L'histoire, en même temps, a produit des modes. D'autres travaux ont laissé plus de traces, et, grâce à leur transformation en outils, sont venus s'insérer dans la palette des instruments utilisés par les évaluateurs du monde entier. L'outil de la comparaison entre des échantillons tirés de façon aléatoire a connu un regain d'intérêt récemment ; il s'intègre bien dans une variété de pratiques qui se réfèrent, dans le *New Public Management*, à la production de « preuves ». Mais si l'on observe l'histoire longue du management public et des *policy sciences*, celle-ci ne démontre pas la focalisation sur un outillage unique, un standard qui s'imposerait scientifiquement dans tous les cas. La pluralité des méthodes et des outils est la règle. En tant que science éclectique, le management, et avec lui sa province l'EPP se présentent plutôt comme des champs d'accumulation de théories partielles et d'instruments en constant renouvellement.

Bibliographie

- BARBIER J.-C., (1982). Heurs et malheurs des libertés en Chine. *Les Temps Modernes* avril, 1773-1837.
- BARBIER J.-C., (1999). article Gestion In ANSART P. et AKOUN A., dir., *Dictionnaire de sociologie*, Seuil, Paris.
- BARBIER J.-C., (2008). La transformation des forums de politique sociale en France depuis les années 1990 In O. Giraud et P. Warin, dir., *Politiques publiques et démocratie*, La Découverte, Paris, 217-239.
- BARBIER J.-C., (2010). Éléments pour une sociologie de l'évaluation des politiques publiques en France. *Revue française des affaires sociales* 1-2, 27-49.
- BARBIER J.-C., (2012). Évaluations « expérimentales », quelques leçons de l'histoire. *Politiques sociales et familiales* 110, 19-32.
- BARBIER J.-C., (2012). Introduction In BARBIER J.-C. and HAWKINS P., eds., 2012, *Evaluation Cultures, Sense-Making in Complex Times*, Transaction Publishers, New Brunswick, 1-18.
- BARBIER J.-C. and HAWKINS P., eds., (2012). *Evaluation Cultures, Sense-Making in Complex Times*. Transaction Publishers, New Brunswick.
- BARBIER J.-C. ET MATYJASIK N., (2010). Évaluation des politiques publiques et quantification en France : des relations ambiguës et contradictoires entre les disciplines. *Revue française de socio-économie* 5, 123-140.
- BEZES P., (2009). *Réinventer l'État, les réformes de l'administration française (1962- 2008)*. PUF, Paris.
- BOUDON R., (2003). *Existe-t-il encore une sociologie ?* Odile Jacob, Paris.
- BOYLE R., G.MCNAMARA, J., O'HARA, (2012). Riding the Celtic Tiger : Forces Shaping Evaluation Culture in Ireland in Good and Bad Times In Barbier J.-C. and Hawkins P., eds., 2012, *Evaluation Cultures, Sense-Making in Complex Times*, Transaction Publishers, New Brunswick, 45-64.
- CAMPBELL D. T., (1969). Reforms as Experiments. *American Psychologist* 24. (4), 409-429.
- DESROSIÈRES A., (2008). *Gouverner par les nombres, L'argument statistique II*. Les Presses de Mines Paris Tech, Paris.
- DUFLO E., BANERJEE A., (2009). L'approche expérimentale en économie du développement. *Revue d'économie politique* 115 (9), 692-726.
- DURAN P., (2010). L'évaluation des politiques publiques : une résistible obligation. *Revue française des affaires sociales* 1-2, 5-24.
- FURUBO J.-E., RIST R., SANDAHL R. (eds), (2002). *International Atlas of Evaluation*. Transaction Publishers, New Brunswick and London.
- GIBERT P., (2008). Un ou quatre managements publics ? *Politiques et management public* 26 (3), 7-23.
- GUBA E., LINCOLN Y. S., (1989). *Fourth-Generation Evaluation*. Sage, Newbury Park.
- HATCHUEL A., (2008). Quel horizon pour les sciences de gestion ? Vers une théorie de l'action collective In DAVID A., HATCHUEL A. et LAUFER R., dir., *Les nouvelles fondations des sciences de gestion*, Vuibert, Paris.
- HECKMAN J., (2010). Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of Economic Literature* 48 (2), 356-398.
- JACOB S., (2005). La volonté des acteurs et le poids des structures dans l'institutionnalisation de l'évaluation des politiques publiques (France, Belgique, Suisse et Pays-Bas). *Revue française de science politique* 55 (5-6), 835-864.
- JOBERT B., (1998). La régulation politique : le point de vue d'un politiste In COMMAILLE J. et JOBERT B., *Les métamorphoses de la régulation politique*, LGDJ, Paris, 119-144.
- KING D., (1999). *In the Name of Liberalism*. Oxford University Press, Oxford.
- LABORIER P., AUDREN F., NAPOLI P., VOGEL J., (2011). *Les sciences caméras, activités pratiques et histoire des dispositifs publics*. PUF, Paris.
- LABROUSSE A., (2010). Nouvelle économie du développement et essais cliniques randomisés : une mise en perspective d'un outil de preuve et de gouvernement. *Revue de la régulation* [en ligne] 1^{er} semestre/Spring 2010, consulté le 11 mai 2013. URL : <http://regulation.revues.org/1818>

- LAUFER R., (2008). Où est passé le management public ? Incertitude, institutions et risques majeurs. *Politiques et management public* 26 (3), 25-48.
- MONNIER E., (1992). *Évaluations de l'action des pouvoirs publics* 2^e éd. Economica, Paris.
- NIOCHE J.-P., (1982). De l'évaluation à l'analyse des politiques publiques. *Revue française de science politique* 32 (1), 32-61.
- PALUMBO D. J., (1987). *The Politics of Program Evaluation*. Sage, London.
- PAWSON R., TILLEY N., (1997). *Realistic Evaluation*. Sage, London.
- PERRET B., 2001, *L'évaluation des politiques publiques*. La Découverte, Paris.
- PRESSMAN J., WILDAVSKY A., (1979). *Implementation* 2nd ed. University of California Press, Berkeley.
- ROGERS P., (2008). Using Programme Theory for Complicated and Complex Programmes. *Evaluation : the International Journal of Theory, Research and Practice* 14 (1), 29-48.
- ROSSI P. H., FREEMAN H. H., (1993). *Evaluation, a Systematic Approach*. Sage, London.
- SOCIÉTÉ FRANÇAISE DE L'ÉVALUATION (SFE), (2011). Évaluation des impacts des programmes et des services publics. *Cahiers de la SFE* 6, mai.
- SOLESBURY W (2001). Evidence Based Policy : whence it Came and where it's Going, ESRC UK Centre for Evidence Based Policy and Practice. *Working Paper 1*, Queen Mary University of London, London.
- SPENLEHAUER V., (1995). *L'évaluation de politique, usages sociaux. Trois études de cas d'évaluations*. L'Harmattan, Paris.
- SPENLEHAUER V., (1999). Intelligence gouvernementale et sciences sociales. *Politix* 12 (48), 95-128.
- SPENLEHAUER V., (1998). *L'évaluation des politiques publiques, avatar de la planification*. Thèse de doctorat de science politique, Grenoble.
- STAME N., (2012). The Culture of Evaluation in Italy In Barbier J.-C. and Hawkins P., eds., *Evaluation Cultures, Sense-Making in Complex Times*, Transaction Publishers, New Brunswick , 19-44.
- THOENIG J.-C., (2008). Politiques publiques et cycles de vie. Le bébé et l'eau du bain. *Politiques et management public* 26 (3), 57-76.
- TOULEMONDE J., (2000). Evaluation Culture(s) in Europe : differences and Convergence between National Policies. *Vierteljahrshefte Zur Wissenschaftsforschung* 3, 350-357.
- VAN DER KNAPP L., LEEUW F., BOGAERTS S. NIJSSEN L., (2008). Combining Campbell Standards and the Realist Evaluation Approach. *American Journal of Evaluation* 29 (1), 48-57.
- VAN RIEPER P., (1983). The American Administrative State, Wilson and the Founders – An Unorthodox View. *Public Administration Review* 43 (6), 477-490.
- WEBER M., (1996). *Sociologie des religions*. Gallimard, Paris.
- WEISS C.H., (1993). Where Politics and Evaluation Research Meet. *Evaluation Practice* 14 (1), 93-106.
- WEISS C.H., MURPHY-GRAHAM E., PETROSINO A., GANDHI A., (2008). The Fairy Godmother – and her warts, making the dream of evidence-based policy come true. *American Journal of Evaluation* 29 (1), 29-47.
- ZARKA Y.-C., (2009). Qu'est ce que tyranniser le savoir ? *Cités*, 37 (1), 3-6.